Mad in America
SCIENCE, PSYCHIATRY AND SOCIAL JUSTICE

Many Results in Psychology and Medicine Are False Positives

Only 27.3% of the results in psychology and only 5.3% of the results in medicine have "strong evidence" for being true.

By **Peter Simons** - May 27, 2024

A recent study finds that the likelihood of true findings in psychology is hugely overestimated. The researchers conclude that the presence of a positive effect dropped from 98.9% to 55.7% after controlling for publication bias. That is, existing meta-analyses in psychology find a positive result almost 100% of the time, but after accounting for publication bias, about half of these can be considered false positives.

The proportion of findings in psychology that have "strong evidence" was even lower—dropping to 27.3%. A little over a third (38.39%) met criteria for "moderate evidence."

> **"Publication selection bias, where studies with significant or positive results are more likely to be reported and published, distorts the available scientific record," the researchers write.**

The study was led by František Bartoš at the University of Amsterdam and Maximilian Maier at University College London, and published in *Research Synthesis Methods.* The authors also included renowned researcher John Ioannidis.

Ioannidis is known for a seminal 2005 paper titled "Why Most Published Research Findings Are False," in which he made a statistical argument that the medical literature is full of false positives. Since that time, this has become known as the "replication crisis" and is considered especially pernicious in the social sciences, particularly psychology. Ioannidis has also critiqued the mass production of biased clinical trials, which other researchers have argued are biased enough to be considered "hijacked for marketing purposes," rather than science. These trials are then included in systematic reviews and meta-analyses. However, Ioannidis writes that since most of the findings included in these analyses are likely false positives to begin with, the conclusions of reviews that include them are also false positives. It's a case of garbage in, garbage out.

Publication bias is sometimes used as a catch-all term that takes several forms, but always involves the selective or biased reporting of results. One aspect is the tendency for journals to publish positive studies and leave negative findings unpublished. In outcome reporting bias, researchers include only the positive results from their study, and leave the negative findings out of the final paper. Similarly, "spin" is when researchers focus on the positive results in their abstract and conclusions, while leaving the negative findings hidden deep in the data, or when researchers misleadingly report negative results as positive.

Many previous studies have identified publication bias, outcome reporting bias, and spin as significant problems in the psychiatric literature. These biases have hidden the harms of treatments, inflated the efficacy of treatments, and overall corrupted the validity of the clinical literature.

One of the most high-profile exposés on the impact of publication bias came in the form of a 2008 paper in the top-tier medical journal, *The New England Journal of Medicine.* In that paper, Erick Turner and his colleagues found that after identifying unpublished studies, 49% of the trials of antidepressants showed that the drugs were no better than placebo. Yet almost all of the negative studies never saw the light of day—and those that did were "spun" to appear positive.

Similar findings have been found for alprazolam (Xanax), for which the FDA notes that only one of the five key trials actually found a positive result. More recently, esketamine nasal spray was approved by the FDA despite failing to beat placebo in five out of six clinical trials.

A Deeper Dive into the Current Study

In the current study, the researchers compared meta-analyses in four different fields: psychology, medicine, environmental science, and economics. In total, they included more than 68,000 meta-analyses, with over 700,000 effect size estimates. They used Bayesian statistics to estimate the strength of evidence for an effect and probability of false positives due to publication bias.

In psychology, environmental science, and economics, almost 100% of the results found in meta-analyses were considered "positive"—which is highly unlikely (it's almost impossible that every single hypothesis in the three social sciences just happened to be true). Medicine was the true outlier here, as only 38% of the results in the published meta-analyses were positive.

The researchers found that economics showed the greatest reduction in likelihood of true effect. The

probability of a true effect decreased from the published 99.9% to 29.7% after accounting for bias. It went down to 19.6% when including only findings with "strong evidence" for a true effect.

As mentioned above, psychology showed a massive drop too, with the probability of a true effect decreasing from the published 98.9% to 55.7% (27.3% with "strong evidence").

Environmental science fared a little better, dropping from 99.8% to 70.7% (30.7% with "strong evidence").

Medical meta-analyses showed the smallest drop in likelihood of an effect—but, again, that's because they started with a very low rate of positive findings (38%, dropping to 29.7%). In total, medical findings actually fared the worst when it came to having strong evidence for an effect: only 5.3% of the findings in medical meta-analyses had "strong evidence" for being true.

The researchers also found that the meta-analyses inflated effect sizes—measuring how powerful an effect (or treatment) is, rather than simply whether or not it exists. They found that in economics, effect sizes were more than twice as large as they should be, after accounting for publication bias. In medicine and environmental science, effect sizes were inflated by more than one and a half times. Psychology actually fared the best in terms of effect size inflation, with their results exaggerated by only 40%.

Ultimately, findings in the social sciences are untrustworthy because almost every single analysis yields a positive result, but less than a third are actually likely to be true positives. Findings in medical science are somewhat more trustworthy, but that's only because most of the findings are already reported accurately as false. Yet only 5.3% of the results from medical meta-analyses actually have strong evidence for being true. And the size of the effect is inflated, too, making treatments seem more impactful than they actually are.

> **"In this paper, we documented the considerable impact of publication selection bias on meta-analyses in a variety of disciplines. Even though we can probe the footprint of these biases with the statistical techniques employed here, science ultimately needs to progress toward mitigating publication bias already while conducting and publishing the research," the researchers write.**
>
> **They add, "Our results suggest that the social sciences might especially benefit from adopting practices to mitigate these, including: preregistration, greater transparency, and registered reports."**

****

Bartoš, F., Maier, M., Wagenmakers, E. J., Nippold, F., Doucouliagos, H., Ioannidis, J. P. A., . . . & Stanley, T. D. (2024). Footprint of publication selection bias on meta-analyses in medicine, environmental sciences, psychology, and economics. *Research Synthesis Methods, 15*(3), 500-511. (Full text)

Peter Simons

Peter Simons was an academic researcher in psychology. Now, as a science writer, he tries to provide the layperson with a view into the sometimes inscrutable world of psychiatric research. As an editor for blogs and personal stories at Mad in America, he prizes the accounts of those with lived experience of the psychiatric system and shares alternatives to the biomedical model.